DOCUMENT RESUME

ED 137 415 TM 006 224

AUTHOR Worthen, Blaine R.
TITLE A Look at the Mosaic of Educational Evaluation and Accountability. Research, Evaluation, and Development Paper Series No. 3.
INSTITUTION Northwest Regional Educational Lab., Portland, Oreg. Office of Research and Evaluation Services.
PUB DATE Aug 74
NOTE 42p.
AVAILABLE FROM Office of Marketing and Dissemination, Northwest Regional Educational Laboratory, 710 S.W. Second Avenue, Portland, Oregon 97204 (Order No. 806-5203, $2.50)

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS Decision Making; Definitions; *Educational Accountability; Educational Legislation; *Evaluation; *Evaluation Methods; Evaluation Needs; Formative Evaluation; *Program Evaluation; Summative Evaluation
IDENTIFIERS Colorado

ABSTRACT

A few major concepts about evaluation and accountability are examined briefly, and related to one another in a manner providing a simple portrayal of part of the mosaic of educational evaluation and accountability. This paper is not intended for evaluation specialists or schoolmen well versed in evaluation theory and practice, but for the practitioner who wants a brief summary of some of the more important notions about evaluation which have been presented during the past several years. Specifically, the following topics are presented: (1) simple illustrations of differences in evaluation, research, assessment, measurement, and accountability; (2) a discussion of some basic evaluation constructs; (3) an analysis of the evaluation features of one accountability law, the 1971 Colorado Accountability Act; and (4) general touchstones for judging the adequacy of an evaluation. (RC)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED137415

Research Evaluation Development

Paper Series

Not to be reproduced without permission

TM006 224

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

2

Office of Research and Evaluation Services

National College of Education Library

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY

E. Tyler

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

A LOOK AT THE MOSAIC OF EDUCATIONAL EVALUATION AND ACCOUNTABILITY

No. 3

Blaine R. Worthen
Northwest Regional Educational Laboratory

August 1974

$2.50

The points of view or opinions expressed in this publication do not necessarily represent the official position or policy of the Northwest Regional Educational Laboratory.

The Research, Evaluation and Development (RED) paper series is published by the Northwest Regional Educational Laboratory (NWREL). It presents papers dealing with both theoretical and applied aspects of research, evaluation or program development. Authors, within or outside of NWREL, wishing to submit articles for possible publication in the RED series may contact the editor, RED Paper Series, at NWREL. Persons wishing to order current or back copies of the RED series may contact the Director of Marketing and Dissemination. A price list for titles in the series is located at the end of this paper.

Northwest Regional Educational Laboratory
710 S.W. Second Avenue
Portland, Oregon 97204

A Look at the Mosaic of Educational Evaluation and Accountability[1]

The terms "evaluation" and "accountability" are becoming so well ingrained in educational parlance that it would be easy to assume that the corresponding activities are well understood by educators and well entrenched in educational practice. Phrases such as "program accountability" and "using evaluation to support decision making" appear more and more frequently in educators' writings and conversations in which they describe school activities.

A closer look reveals that what is observed may be largely a form of semantic orthodoxy. Demands to make educational systems accountable to their publics are proliferating at a rapid pace. Yet, as Glass (1972) has noted, most of the activities which masquerade as forms of accountability fail to result in real accountability. More and more legislative bodies are authorizing funds for the express purpose of evaluating educational programs to determine their effectiveness. Yet many of the resulting systems fail to hold the schools accountable at all and deteriorate into mandated information management or testing systems which add little if anything to the quality of education. Verbal statements about evaluation and accountability are abundant, but genuine evaluation of educational programs is infrequent.

One reason for the scarcity of good examples of evaluation and accountability in education probably lies in the fact that school practitioners

---

[1] This paper is based on the author's script for two programs on evaluation in a 1973 University of Colorado state-wide educational television series entitled "Educational Accountability."

have had little usable guidance in how to facilitate or conduct such activities. The evaluation literature is badly fragmented into unrelated pieces and is as difficult to synthesize as it is to make a meaningful picture from a random handful of pieces to a jigsaw puzzle. Looking at the individual pieces is little more helpful, for the level of discourse in individual writings is often aimed at fellow evaluation theorists more than at schoolmen, thereby communicating a great deal of detail about a topic which lacks a larger context within which it could be useful. Working under this handicap, busy practitioners can hardly be faulted for not expending the necessary time to try to develop a clear picture from the current evaluation literature.

The purpose of this paper is to examine briefly a few major concepts about evaluation and accountability and relate them to one another in a way that will provide a simple portrayal of part of the mosaic of educational evaluation and accountability. Two caveats should be stated at the outset. First, this paper is not intended for evaluation specialists or schoolmen well versed in evaluation theory and practice. It is primarily intended for the practitioner who wants a brief summary of some of the more important notions about evaluation which have been presented during the past several years. Second, the basic thread which will run through this paper is evaluation, with accountability playing only a supporting, illustrative role.

An Attempt at Definition

Evaluation is closely related to several other terms with which it is often associated and generally confused--terms like research, assessment, measurement, and, of course, accountability. These terms should be separated from one another, since the meaning we attach to words often

influences what we do (Glass and Worthen, 1971). It is not my intent here to engage in the usual academic activity of defining one term by use of another that is equally arcane. Those who find dictionary style definitions helpful should read earlier writings by Wardrop (1972) or Worthen and Sanders (1973). I would prefer to use some very simple-minded examples to illustrate differences in five interrelated but different concepts: measurement, assessment, evaluation, accountability, and research.

A high school invitational pole vaulting meet, in which a number of boys participate, can serve as an example. The performance of the boys could be viewed in relation to each of the five concepts, as illustrated below.

Measurement answers the question, "How high did each boy vault successfully?" It is the simple act of determining the maximum height at which each boy cleared the bar.

Assessment answers the question, "How well did each boy meet the goal or objective he (or his coach) set for him?" Assessment comprises three activities: (1) decisions about goals or objectives; (2) measurement of how well objectives are attained; and (3) a summary of the measurement information in relation to the objectives or to relative performance. Pursuing our example, a minimum objective is reflected in the decision to set the bar at 10 feet for the initial round since it is expected all the boys can clear that height. Individual objectives are reflected in decisions to try to exceed a height of 15 feet or break the record of 16 feet. Decisions about how to measure attainment of objectives are evidenced in established rules that height of jumping should be measured in feet and inches and a miss occurs when a boy knocks the bar off three consecutive times at a given

height. Measurement occurs when those rules are applied. The statement that all of the boys cleared 10 feet but only two of them were able to clear 15 feet is a brief summary of the measurement information in relation to the objectives or to relative performance.

Evaluation answers questions like "Given a standard such as height, which boy is the best pole vaulter?" "Overall, did the use of bamboo poles or steel poles result in greater heights in vaulting?" "Which type of pole broke most often during a vault?" "Did the training program used by a particular high school produce satisfactory results?" Evaluation includes (1) determining what measures and standards should be used to judge performance (e.g., height of highest successful vault, consistency of successes without a miss, form), (2) deciding whether the standard should be relative (e.g., compared to other boys) or absolute (e.g., a state-wide minimum height for qualifying), (3) collecting the relevant information through measurement or other means, and (4) applying the standard in determining merit or effectiveness.

Accountability answers the question "Were the coaches and athletic programs responsible for preparing the boys for the meet successful in helping their boys reach expected performance levels and/or win the meet?"

Research answers questions like "What are the characteristics of steel poles or cross-handed grips which make them superior to their counterparts?" "Why does athletic program A produce better results than program B?" In the pole vaulting example, the primary function of research would be to determine why certain performance levels were reached.

These examples oversimplify the five concepts but may help to illustrate major differences among them and reveal why my focus in this paper will be primarily on evaluation, secondarily on accountability, and not at all on research, assessment, or measurement. Research is clearly an enormously complex undertaking which goes far beyond simple evaluative findings (e.g., Program A is better than Program B on a particular criterion) to try to fix the causes for those findings. The complex activities inherent in such pursuit of causal explanations makes genuine research a luxury few school districts can afford. Assessment has many of the trappings of evaluation and shares with it many common activities, but it lacks evaluation's explicit judgments of worth or effectiveness. Assessment generally is used to depict something in detail, looking at it through a frame established by the goals or objectives, but it stops short of judging whether the resulting portrait is good or bad, tasteful or tasteless. Useful as assessment is, going beyond it to a complete evaluation of an educational program is critical to attempts to improve school programs. Measurement is simply a process for collecting the data on which evaluative judgments will be made. It is a key tool in evaluation but hardly suffices in and of itself. Accountability is a broad concept which goes beyond evaluation but obviously depends on evaluation as one of its central steps, making a discussion of evaluation essential to any discussion of accountability.[2]

---

[2] Glass, in a statement quoted later in this paper, argues that evaluation is not an essential ingredient in accountability. In the broad and relatively pure type of accountability he describes in his writings (Glass, 1972), the argument is valid. However, I fail to see much utility in a system which merely served to disclose performance without also providing standards by which the performance could be judged.

If this attempt to sort evaluation out from near "look-alike" terms leaves some of the distinctions blurred, they hopefully will become more clear as the concepts of evaluation and accountability are discussed in the remainder of this paper.

Several Views of Evaluation

Until 1965, the term evaluation was generally used in education as a synonym for grading. Little real evaluation of educational programs per se had taken place.[3] With the passage of the Elementary and Secondary Education Act and its accompanying mandate that all Title I and II projects must be evaluated, a more general concern for evaluation was registered--in some ways a preview of the accountability movement to come. During this period, many prominent methodologists and educationists turned their attention to how educational programs should be evaluated. Many evaluation "models" emerged, ranging from near prescriptions for how evaluations should be carried out to presentations of a few factors which should be considered in any evaluation. These models have appeared in the literature and, in the absence of a good empirical base for determining the best way to evaluate educational programs, have greatly influenced the present practice of evaluation.[4] These models have been reviewed elsewhere (Worthen and Sanders, 1973) and it is not my intent to summarize them here. Instead,

---

[3] Obviously, there are notable exceptions to this statement such as the Eight Year Study, but they are clearly the exception rather than the rule.

[4] The impact and general shortcomings of the models has been discussed previously (Worthen, 1972) and will not be repeated here.

I would like to quote statements made by some of these leading thinkers in evaluation which summarize their views of what evaluation is and provide a backdrop against which to discuss evaluation and accountability in further detail.[5]

Ralph W. Tyler

> Educational evaluation is finding out what students have learned in their school work and which of them are having difficulty in learning. For example, in the primary grades how many children have learned to read, to add, subtract, multiply and divide, and to work cooperatively with other children? Which ones are having difficulty in learning these things? Have they learned other things of value? Or, as another example, in the high school how many youth have learned to write clearly, understand the basic principles of our Constitution, and can explain the processes of Nature? Have they learned other things of value in the high school? Which are having difficulty? All of us, whether in education, business, health services, or other fields need to know how we are doing. Are we really attaining our purposes and to what extent? Are we having difficulties? What are they? Are there improvements that need to be made? Educational evaluation is important and necessary both to help the teacher and to give the public a better notion of our educational achievements and where our problems lie that require thoughtful attention. (Tyler, 1973).

[5] The statements quoted herein are taken from audiotaped statements originally included in the television program referenced earlier. The charge to the persons quoted was to prepare a brief (one to three minute) statement completing the phrase "Evaluation is . . . . " for a practitioner audience. Readers should keep in mind the severity of the constraints imposed by that charge.

W. James Popham

When people evaluate, they make an appraisal of some kind. They make an estimate or judgment of the worth of some phenomenon, and in educational evaluation we are concerned with making appraisals of the worth of educational enterprises. The major activities of educational evaluators can be focused in three general arenas. Two of these have to do with specification of the intentions we want to accomplish in our educational endeavors. Many people refer to these as "statements of instructional objectives."

The first kind of evaluation we have to engage in attends to the objectives themselves. Which objectives are really worthwhile pursuing? Which are worth accomplishing even if we could? To evaluate educational objectives, we discover that the more precisely they are articulated, the more rationally we can decide upon which objectives should or should not be pursued in our schools. We are beginning to devise ways whereby students, teachers, scholars, citizens, everyone who has a stake in the educational enterprise, can appraise the worth of educational objectives. The more precisely those objectives are explicated, the better the evaluative judgments can be.

A second focus of educational evaluators concerns the assessment of the degree to which objectives have been achieved. Once we decide upon the really worthwhile goals, a second task is to discern whether the objectives have been realized as a consequence of our educational endeavors. Once more we find that an explicitly stated objective, stated in terms of measurable learner behaviors, permits more readily such assessment. We can discover whether such objectives have been achieved. And educational evaluators are very much concerned with discovering the degree to which objectives have been achieved.

A third focus of educational evaluation these days is upon judging all of the effects of instructional endeavors, both those which were intended (as reflected by the objectives) and those which were not anticipated at all. In other words, rather than being attentive only to what intentions the instructional designers had at the beginning of instruction, we should attend to all the results of an instructional endeavor, those that were anticipated as well as those which were unforeseen. (Popham, 1973).

Robert E. Stake

Let us look more carefully at the notion of evaluation. To me, it is mostly a matter of saying something is good or bad, or saying how good it is or how bad it is. In order to communicate effectively with other people when evaluating, we have to talk about what it is that we are evaluating, and that may take a great many words and a few pictures. It may take many different displays to indicate to other

people what it is that you are evaluating and it may take a number of different judgments of goodness or badness, judgments from different people, in order to give the person we are talking to some confidence that we have found an answer to the value of the program. (Stake, 1973).[6]

Daniel L. Stufflebeam

What is evaluation? What is it for? What questions does it address? Who should do it? How should they do it? And by what standards should their work be judged? Persons' responses to these questions can reveal whether they have thought very much about evaluation and if so, what their conceptualizations are. Briefly, here are my responses to these six questions.

1. Evaluation is the ascertainment of merit.

2. Evaluation serves both decision making and accountability.

3. Evaluation addresses questions about goals, designs, procedures, and results.

4. Evaluation for decision making can appropriately be performed by an agency staff, but external personnel should be involved in evaluation for accountability.

5. The process of doing evaluation involves delineating the evaluation requirements, obtaining the relevant information, and providing the obtained information to the appropriate audiences.

6. Evaluative information should be judged for its technical adequacy, its utility, and its worth compared to its cost.

As a final point, evaluation should serve not only to prove the worth of programs, but also to improve them. (Stufflebeam, 1973).

---

[6] Stake's statement was excerpted from an audiotape, "The Teacher and Accountability," produced by the VIMCET Corporation and reproduced here with their permission.

Michael Scriven

Evaluation is the systematic and objective determination of the merit or worth of something. In the educational field, of course, this something is usually an educational product or process--or perhaps educational personnel. But having a definition tells you little about how to do it. The way I conceptualize doing it, a very heavy stress should be placed on the comparative element in evaluation. I think evaluation is very rarely of any interest unless it tells you something about how well the thing you are looking at did by comparison with other things that are available or could be set up at comparable cost. So, to me the main task in educational evaluation is identifying the most important comparisons to be made--the critical competitors--and then proceeding to document the comparisons on the various dimensions of interest to the respective audiences of the evaluation. Correctly done, this approach avoids a very serious flaw in a great deal of educational evaluation, that of regarding everything as necessarily to be evaluated in terms of the goals of its developer or its designer.

To sum it up, the evaluation approach which will be most fair to educational practice is (1) the constant juxtaposition of the item to be evaluated with various critical competitors and (2) the critical analysis of the dimensions of difference of performance between these with respect to the needs of the target population rather than with respect to the goals of the producer. That's evaluation. (Scriven, 1973).

Gene V Glass

Evaluation is the assessment of the worth or value of a thing. But rather than talking about what evaluation is, I want to say something about what evaluation is not. Think of the word BOATS as a mnemonic reminder of the things that look like evaluation, but really are not.

Budgeting. Budgets are useful. Every school district should have one and probably does. Somehow, perhaps because Program Planning and Budgeting was mandated along with accountability in Colorado, drawing up budgets has become confused with being accountable and, worse yet, with evaluating school programs. Budgets can be based on nothing more than whims, fads, and poor judgment. When you get right down to it, a superintendent drawing up a budget is not necessarily evaluating or judging anything. He is only drawing up a budget.

Objectives. Objectives, goals, and intentions are basically the same things. Stating objectives is sometimes the first step in evaluation, sometimes not. Under some circumstances the evaluator need not concern himself with the program objectives at all. After all, intentions are only intentions. It is the value of what eventuates from a school program that counts.

Accountability. In my opinion, the principle of disclosure is at the heart of true accountability. The accountable school discloses its goals, decision making procedures, financial affairs, and its accomplishments, good and bad. I would regard a school as acting accountably if it

merely disclosed such facts to the public, even if it could not accomplish the more difficult job of turning the facts into value claims. It is not necessary to evaluate to be accountable.

Testing. Millions of standardized tests are given every year in schools and, on balance, I suspect they are worth the cost. Tests are often used in program evaluation, but there are many steps between testing children and validly judging the worth of the programs they take part in. The methodology of evaluation is a guide for use in moving from evidence, including test scores, to judgments of value. The technology of testing has little to do with deriving value judgments, so although your school may administer a great many tests, it does not follow that it is doing a lot of evaluation.

As for the S in BOATS, I prefer to let it dangle on the end. I would not want things to work out too neatly. They never do in reality, neither in teaching nor evaluation. (Glass, 1973).

As the statements above show, their authors differ somewhat in their views of what evaluation is and how it should be carried out, making it impossible to combine their notions into any single prescription for how to evaluate a particular school program. However, there are some common issues to which each of these evaluators have attended in one way or another and their divergent suggestions provide a set of alternatives from which practitioners can select in conducting a program evaluation. In the remainder of this paper, I will present a few simple concepts about evaluation (and to a lesser extent accountability) and discuss the alternatives which exist for schoolmen as they approach each concept.

Some Basic Evaluation Concepts

The first topic I would like to discuss is the relationship of evaluation to decision making in the schools.

Evaluation for decision making. The basic notion here is very simple and has been well articulated by Stufflebeam in his writings (e.g., Stufflebeam, et al., .1971). Put simply, the idea is that evaluation exists to provide information to administrators so they can make more intelligent decisions about the programs they administer. Administrators obviously must make decisions about educational programs, whether or not they have adequate information about which of the alternatives they are choosing from is best for their purpose. The role of evaluation as it relates to decision making is to examine each alternative critically and make a judgment about its worth for the purpose the administrator has in mind. Such collection of evaluative data to help make intelligent decisions is the major use for evaluation as I see it relating to educational programs.

Evaluation of decision making. This is the process of looking explicitly at how the administrator goes about reaching a decision. Did he consider alternatives? Did he look at all the data? Was he unduly influenced by political considerations? This is a specialized use of evaluation where decision making is merely the object of the evaluation--an interesting and important use of evaluation but not the major one of concern in this paper.

Evaluation of the impact or results of decision making. Accountability legislation in several states has led to this use of evaluation. Given that a particular decision has been made, what impact has that decision had on the quality of educational programs? The authors of the 1971 Colorado accountability act evidently had this in mind when they stated that it was necessary to establish a ". . . means for determining whether decisions affecting the educational process are advancing or impeding student achievement" (Colorado

Senate Bill No. 33, p. 2).[7] It seems strange that they would press to evaluate the impact of decisions without first seeing that the decision maker has good evaluation data--and uses it. This suggests a certain innocence on the part of the legislators about how to really effect educational improvements. It is not very useful to worry only about whether the decisions being made are helpful, hurtful, or of no consequence when equal concern should be shown for how to improve those decisions. Legislatures and school systems should not become so preoccupied with the outcomes of the decision making process that they fail to solve problems those outcomes might reveal. It seems advisable in this context to use evaluation as Stufflebeam has proposed, as a mechanism for administrators' use in judging decision alternatives to help them make better informed decisions. If evaluation is used effectively in this way, looking at the results of decision making should reveal fewer decisions that have affected educational programs negatively or not at all.

Beyond its relationship to decision making, there are two additional dimensions of evaluation which are interrelated and should be discussed together. These are formative vs. summative evaluation and internal vs. external evaluation.

Formative and Summative Evaluation

Scriven (1967) first distinguished between formative and summative evaluation. Since then, the terms have become almost universal in their

---

[7] The 1971 Colorado Accountability Act is discussed throughout this paper as an example of public school accountability legislation. However, because of basic similarities in the Colorado law and accountability legislation in other states, the discussion herein can be generalized to many of the other states where school personnel are faced with the task of implementing new accountability laws.

use in the field. Although in practice distinctions between these two types of evaluation may blur somewhat, since they are not strictly orthogonal, it seems useful to summarize the major differences noted by Scriven, even at the risk of some oversimplification.

Formative evaluation simply refers to evaluation that is conducted during the operation of a program for the express purpose of providing evaluative information to program directors for their use in improving the program.[8] For example, during the development of a curriculum package, formative evaluation would involve content inspection by experts, pilot tests with small numbers of children, field tests with larger numbers of children and teachers in several schools, and so forth. Each of these steps would result in immediate feedback to the developers who would use the information to make necessary revisions in the materials.

Summative evaluation is evaluation conducted at the end of a program for the express purpose of judging the worth or effectiveness of that program for potential users for whom it has been developed. For example, after the curriculum package is completely developed, a summative evaluation might be conducted to determine how effective the package is with a national sample of typical schools, teachers, and students at the level for which it was developed. Note that the audience here is very different. In formative evaluation, the audience for the evaluation report comprises personnel in the program--in our example, those who were responsible for developing the curriculum package. In summative evaluation, the audiences for the evaluation report include the potential users (students, teachers, and other professionals) and the source of funding (taxpayer or funding agency), as well as program personnel.

[8] The discussion in the remainder of this paper is intended to apply equally to evaluation of educational programs, projects, products, and processes--indeed, any object of an educational evaluation. However, to avoid tedious redundancy, only one term (e.g., "program") will generally be used hereafter in each example or concept presented. The other possible objects of educational evaluation can be assumed to be included by implication.

Program development decisions and accountability decisions draw respectively on formative and summative evaluation. Formative evaluation leads to (or should lead to) decisions about program development (including modification, termination, continuation, and the like). Summative evaluation is one of the necessary steps in making accountability decisions. The 1972 Colorado accountability legislation (and many other state accountability laws as well) deals primarily with summative evaluation and emphasizes formative evaluation little if at all. This is unfortunate, not because summative evaluation is unimportant--no right-thinking educator could take that stand--but because without formative evaluation it is incomplete and inefficient. Consider the foolishness of developing a new design for an aircraft and submitting it to a "summative" test flight without first testing it in the windtunnel. The probable success of premature summative evaluations in education seems little greater.

Internal and External Evaluation

The dichotomy of internal vs. external evaluation is largely self-explanatory. The adjectives refer to whether the evaluator is internal (i.e., an employee of) or external to the program being evaluated. A Title III program might be evaluated by an evaluator who is a member of the project staff (internal) or by a site visit team appointed by the State Department of Education (external). There are obvious advantages and disadvantages with both of these roles. The internal staff evaluator is almost certain to know more about the project than is possible for any outsider, but he may also be so close to the project that he is unable to be completely objective in his view of it. There is seldom as much reason to question the objectivity of

the external evaluator (unless he is found to have a particular ax to grind) and this dispassionate perspective is perhaps his greatest asset. Conversely, it is difficult for an external evaluator to ever learn as much about the project as the insider knows. Note that when I say as much, I refer only to quantity, not quality. One often finds an internal evaluator who is full of unimportant details about the project but overlooks several critical variables. If these bits of key information are picked up by the external evaluator, as is sometimes the case, he may end up knowing much less overall about the project but knowing much more of importance.

Possible Role Combinations

The dimensions of formative and summative evaluation can be combined with the dimensions of internal and external evaluation to form the two-by-two matrix shown in Figure 1.

	Internal	External
Formative	1	2
Summative	3	4

Figure 1: Combination of Evaluation Roles

The most common roles in evaluation might be indicated by cells one and four in the matrix. Formative evaluation is typically conducted by an internal evaluator. His knowledge of the program is of great value here and possible lack of objectivity is not nearly the problem it would be in a summative evaluation. Summative evaluations are typically (and probably best) conducted by

external evaluators. It is difficult, for example, to know how much credibility to accord a SRA evaluation which concludes that a set of SRA reading materials is far better than its competitors.

Another important role--that of the external formative evaluator shown in cell two--is almost completely neglected in educational evaluations. As implied earlier, the internal evaluator may share many of the perspectives and blindspots of the rest of the program staff and, consequently, neglect even to entertain some negative questions about the program. The external evaluator doesn't have long familiarity with the program to fall back on and he is much less likely to be influenced by a priori perceptions that it is basically good. This is not synonomous with saying that he is predisposed toward judging the program as bad. His orientation should be neither positive nor negative, only neutral and uninfluenced by close associations either with the program or its competitors. In essence, the external formative evaluator introduces a cold, hard look of reality into the evaluation relatively early--in a sense a preview of what a summative evaluator might say. This fresh outside perspective is important, even if used infrequently, to avoid the frequent disaster that occurs when a program staff carefully and self-consciously conducts a formative evaluation of their own program, using criteria and variables they interpret as proving the program successful, only to have an outside agency (a school board or site visit team) close the program down because their summative evaluation focused on other variables or used different criteria which resulted in overall negative outcomes. Wisdom would dictate the use of an outside formative evaluator as part of the formative evaluation of every program or product.

Cell three in Figure 1, the internal summative evaluator, strikes me as a role that is only infrequently appropriate. As stated earlier, the summative evaluation is generally best conducted by an external evaluator or agency. However, in some instances there is simply no possibility of obtaining such external help because of financial constraints or absence of competent personnel willing to do the job. In these cases, the summative evaluation is weakened by the lack of outside perspective, but it might be possible to retain adequate objectivity and credibility by choosing the internal summative evaluator from among those who are some distance removed from the actual development of the program or product being evaluated.

The concepts discussed so far lead to consideration of two radically different approaches to program evaluation--goal-directed evaluation and goal-free evaluation. Each of those approaches is described briefly below.

Goal-Directed Evaluation

Goal-Directed Evaluation is perhaps the most common type of evaluation practiced in education. It is basically the approach first suggested by Ralph Tyler as early as the 1930s and reiterated in his statement quoted earlier. This approach has been adopted and expanded by the numerous advocates of the use of behavioral objectives. In essence, it depends on six basic steps:

1. Establishing broad goals for the program.
2. For each broad goal, identifying specific objectives which, if attained, would result in attainment of the goal.
3. Stating each objective in measurable terms.
4. Developing or selecting measures of performance (usually pupil performance) for each objective.

5. Conducting the program which is to attain the objectives.

6. At the end of the program, measuring performance on each objective to see if expected outcomes have been achieved.

The third step provides the genesis for behavioral objectives. As an aside to the discussion of goal-directed evaluation, I am uncomfortable with the current fanaticism about behavioral objectives which seems to have permeated the field of education. It is obviously time that educational goals and objectives are made more explicit and observable and, as Popham stressed in his statement quoted earlier, objectives obviously should be stated in terms in which they can be assessed. But one cannot help being distressed by the mindlessness running rampant through education which would have all educators state every intent--however trivial--in behavioral terms. In some schools, the staffs are spending so much time and energy stating everything they want to teach in behavioral terms that they hardly have time to teach. I am frankly unsympathetic with the zealous efforts to train every teacher to use a recipe to translate every aspiration into a behavioral objective. This is especially true where teachers are used to write objectives intended more for evaluation than for instructional purposes. It is, after all, the evaluator who is supposedly skilled in the language of operationalization. I think the evaluator should take the following stance in working with program personnel: "Give me an objective in any form, just so I understand what your intent is. As an evaluator, I will translate your objective into behavioral terms and have you review my statement to make certain I have not distorted your intent." That makes more sense to me than trying to train all educators to be evaluators. The pendulum obviously needed to move from the irresponsibly soft-headed position that educators do not need objectives because, after all,

they "know in their hearts they are right." But education swung too far to the other extreme when it spawned the religion of behaviorism and the zealots who apply it unintelligently. One can hardly oppose operationalizing instructional objectives and assessing their attainment, but the level of reduction and the utility of scores or hundreds of objectives for each area of endeavor should be questioned. It is simply far too much of a good thing. If not tempered with reason, the press for behavioral reductionism seems likely to backfire by disenchanting educators with all objectives--a result which would cripple educational evaluations.

Goal-Free Evaluation

Goal-Free Evaluation has been recently introduced to the field of evaluation by Scriven (1972). The rationale for goal-free evaluation can be summarized briefly as follows. First, educational goals should not be taken as given; they like anything else should be evaluated. Further, goals are generally little more than rhetoric which seldom reveals the real objectives of the project. In addition, many important outcomes of a program do not fall in the category of goals or objectives anyway (e.g., a Title III project will create additional jobs--a desirable outcome, but never an explicit goal of a Title III project). The most important reason for proposing goal-free evaluation, however, is the salutary effect it will have on reducing bias and increasing objectivity in evaluation. In goal-directed evaluation, an evaluator who is told the goals of the project is immediately limited in his perceptions--the goals provide a set of blinders which causes him to miss important outcomes of the program which are not directly related to the goals (side effects, as they are known in medical parlance). For example, an evaluator who is told that the goals of a dropout

rehabilitation program are to (1) bring dropouts back into school, (2) train them in productive vocations, and (3) place them in stable jobs may spend all of his time designing and applying measures to look at things such as how many dropouts have been recruited back into school, how many have been placed and continued in paying jobs, and so forth. All to the good--and the program may be successful on all these counts. But what about the fact that the crime rate of other (non-dropout) children in the high school has trebled since the dropouts were brought back into the school? Indeed, a hidden curriculum in stripping cars seems to have sprung up! This is a negative side effect which is much more likely to be picked up by the goal-free evaluator than by the goal-directed evaluator who has his built-in blinders imposed by his knowledge of the objectives.

Such a brief summary hardly does justice to Scriven's concept of goal-free evaluation, but it at least provides an introduction to an interesting new approach that is getting a lot of attention in the field of evaluation.

It might be helpful to point out that goal-directed and goal-free evaluation are not mutually exclusive activities. Indeed, they supplement one another very well. The internal staff evaluator of necessity conducts a goal-directed evaluation. He can hardly hope to avoid knowing the goals of the program and it would be unwise to ignore them even if he could. Program directors obviously need to know how well the program is meeting its goals, and the internal evaluator uses goal-directed evaluation to provide him with that information. At the same time, it is important to know how others judge the program, not on the basis of how well it does what is is supposed to do--but on the basis of what it does in all areas, on all its outcomes, intended or not. This is the task for the external goal-free evaluator who knows nothing of the program goals. So it isn't either-or. Both goal-directed evaluation

and goal-free evaluation can work well together. Even Scriven agrees that the major share of the evaluation resources should go to goal-directed formative and summative evaluations. What is tragic is when all the resources go to goal-directed evaluation on a program where the goals do not even begin to include the important outcomes of the program.

Comparative vs. Non-Comparative Evaluation

Another consideration in evaluation which has some relevance to the preceding discussion is that of comparative evaluation vs. non-comparative (or single program) evaluation. There is a long literature on this topic which could not even be listed, let alone summarized here, but the issue can be encapsulated briefly as follows. Comparative evaluations are those where two or more programs or methods are compared with one another on common criteria. For example, assume a public school system is planning to establish an elementary language program in Spanish. All but two sets of curriculum materials have been excluded on the basis of considerations such as cost, guaranteed availability of replacement materials, and the like. In addition to the two sets of printed student and teacher materials, some of the teachers have expressed enthusiasm for a new conversational approach to teaching Spanish which uses no written materials for students but depends exclusively on in-class conversations. A comparative evaluation might be designed to involve a random sample of six elementary schools in the district, with each of the three approaches randomly assigned to two of those schools for use there as the exclusive treatment. The outcomes of the three curriculum approaches could be compared on criteria such as students' conversational ability in Spanish and students' ability to read Spanish, and a judgment made as to which approach is best on these criteria. Obviously, this example

is oversimplified since it ignores treatment-aptitude interactions, weighting of criteria, and the like, but it should serve to illustrate the point.

Non-comparative or single-program evaluations obviously lack any comparison group. The focus of these evaluations is internal and generally is built on a goal-directed approach. Single-program evaluations are the most common type of evaluation conducted in education today. In the previous example, the school system would make a decision on some relevant basis (e.g., reputation of publisher or cost) to try a particular approach to teaching Spanish. Objectives for the program would be carefully noted, the program would be implemented, and, after it had run its course, measures would be applied to see if it had attained its objectives. In short, the basis for judging success would lie not in comparing the program with any other, but in an internal check for discrepancies between what the program purports to do and what it really does.

Both the comparative and single-program evaluation paradigms are well entrenched in education and there is no unequivocal answer as to which is best for all evaluations. It obviously depends on the questions to be answered and the resources available for the evaluation, to mention only some of the determinants. For example, if one is evaluating three house plans offered by a particular builder, the problem can be approached as a single program evaluation. Assume one criterion is "convenience," and an examination of Plan A revealed that the single inconvenient feature was the necessity of crossing an open hallway to carry food from the kitchen to the dining room. So far, so good. Plan A comes very close overall to meeting its goal of convenience. But should one choose it? That all depends on the other two

plans in the same price range. Does it have any bearing on the decision if we find that Plan B has no inconvenient features and Plan C requires that one go through the bathroom to reach the kitchen? Of course, as I can testify from living briefly in a rental unit with the Plan C feature. The point of this example is that without looking at Plans B and C, one would never know what he had selected (or rejected) with Plan A. For this reason, I tend to view comparative evaluation as the ultimate in evaluating educational programs, since it allows you not only to know what you gain by choosing a particular program or method, but also what you give up by rejecting other alternatives. Numerous administrators think they want to know whether a particular program does any good. What they really should be asking is what benefit the program produces, at what cost, and compared with the benefits produced by other alternatives with similar costs. Obviously, there are numerous occasions when comparative evaluations cannot be conducted or are irrelevant to the questions posed. Unfortunately, many appropriate opportunities to conduct comparative evaluations are lost because many educators view "comparative experiments" as useless or even harmful. This perception probably stems from instances where they have seen comparisons conducted unintelligently by those who evidenced in their designs sophomoric misunderstandings of the methodology and its appropriate application. Hopefully this perception can be changed as evaluators learn when and how to structure alternatives in ways that can be demonstrated to have utility for decision makers.

An Analysis of Evaluation Characteristics of One Accountability Law

Enough abstract concepts have been presented that it may be helpful to examine them in a real context. Describing and analyzing the evaluation characteristics of one accountability law, the 1971 Colorado Accountability Act, may be instructive, especially so since accountability legislation in many other states contain similar characteristics.

I suggested earlier that the Colorado legislation includes a very narrow view of how evaluation information might be used in decision making and noted that looking only at the impact of the decision was akin to locking the barn after the horse had escaped. It is unfortunate the focus of the law was not at least as great on how evaluation could be used to improve the quality of decisions and, in turn, their utility.

In writing accountability legislation, many lawmakers seem unclear as to whether their real interest is simply in disclosure, summative evaluation, formative evaluation, or some combination. The activities mandated by the Colorado Accountability Act are primarily summative. On the surface, this seems eminently reasonable, since accounting for the benefits derived from large expenditures of public funds is essentially a summative activity. However, the language of the act and the discussions surrounding its passage make it clear that the legislators were also interested in forcing educators to collect information for immediate use in improving the quality of education. Yet, there is no real provision for the regular use of evaluative data to improve a school's program except at the end of each annual cycle. That tardy schedule is hardly congruent with the intent of providing feedback for program improvement. No one would deny the need for summative evaluation

information about Colorado schools, but it is unfortunate that the legislation fails to place any explicit emphasis on formative uses of evaluation for program improvement since this was one apparent interest of the legislators.

A more serious problem is that the summative evaluations which will result from the application of the Colorado formula will not be very satisfactory either. The law emphasizes internal evaluation within each district and ignores external evaluation of programs in those districts. This would be more understandable if the act focused on formative evaluation. Earlier in this paper, internal summative evaluation (cell three in Figure 1) was presented as the weakest case and the most difficult to implement without bias. Yet this is the cell in which the Colorado law seems to fit best. Probably few lawmakers would condone the practice of asking banks to conduct their own audits, or asking pharmaceutical companies whether one of their drugs should go on the market. Yet the Colorado assembly has asked its schools to conduct the summative evaluations of how well they are living up to the promises they made and to report their conclusions back to the legislature, with the clear implication that the information will influence future allocation of funds. Perhaps the lack of a profit motive in public education is a telling factor, or perhaps educators are simply more trustworthy than bankers and pharmaceutical researchers. Conversely, one could speculate that educators have no special moral eminence, especially when faced with accountability mandates which many perceive as unreasonable and punitive. In such a context, it may be grossly unfair to educators to ask them to carry sole responsibility for evaluating their attainments. Inclusion of external summative evaluators would largely eliminate the conflict of interest in which many Colorado schoolmen feel they have been placed by the accountability act.

The present Colorado law is also a classic example of goal-directed evaluation. The General Assembly mandated that each school district should report once a year to its local constituency and to the state " . . . the extent to which the district has achieved its stated goals and objectives" (Senate Bill No. 33, p. 4). Apparently all a district has to do is state some general goals and specific objectives, carry on their program for a year, and at the end of that time report how well it has done on those goals and objectives. Although it often is important to know whether or not a district attained its stated objectives, such is not always the case. It depends largely on the prior question of whether the goals were worth attaining in the first place. Some goals that are attainable are hardly worth the effort. A more serious problem occurs when all goals were attained not because the program was effective but because the goals were set too low or had already been attained in part through other means. What in the Colorado law is to keep insecure districts from setting goals safely low, or overly ambitious districts from setting goals impossibly high? The first could be applauded and the second censured by lawmakers without the quality or effectiveness of their educational programs really entering into the judgments. This is not to argue against local goal setting, but only to point out that statewide accountability systems might better depend on assessing and disclosing outcomes on "minimum essentials" which should operate in all schools than asking each district to develop local goals and measure their attainment in a way which defers meaningful interpretation of the results. The situation is almost analogous to that in which one needs to identify which children in a classroom are in good health and which are suffering from malnutrition, and height is considered a relevant indicator.

There would be at least a measure of foolishness in asking each child to make his own tape measure, use it in measuring his height, and then report how well he has attained the height he desires to reach at that point or whether he is too tall or too short for his age.

If it is not already patently clear, I am not enamored with the Colorado accountability act as a model of accountability legislation. Even if a district followed it in every detail and specification, the resulting system would fail to qualify as either a good evaluation or accountability system. Perhaps one cannot really expect a legislative assembly to write adequate technical legislation in education and should not be discouraged by such failures. It would seem more productive to focus on the obvious intent of the act. The Colorado law is clearly intended to force school districts to think about and articulate what they want to do and to assess the effectiveness of what they attempt. The General Assembly obviously wants schoolmen to look at where their decisions lead them and try to improve schooling as a result. Rather than criticizing legislators because they exhibit some naiveté about evaluation (an innocence shared by many persons in education), educators could better fulfill their role as responsible professionals by attempting to implement the intent of the legislation. To do so would serve in the best interests of each school district, especially if educators saw the advantages of "piggy-backing" the development of a sound evaluation system onto the need to meet legal requirements. Considerable time will be demanded on the part of schoolmen to meet the minimum "accountability" requirements, and the result could still be an inadequate evaluation system. With some refocusing and a modest increase in time spent, a fully functioning

evaluation system could be developed. Schools could profit greatly if the impetus provided by the legislation could be used as an opportunity to develop a good evaluation system, even though it doubtlessly means exceeding the minimum essentials described by law.

Characteristics of Good Evaluations

What I have presented so far implies that there are good evaluation systems and bad evaluation systems and touchstones to enable educators to tell one from the other. There are some basic components which in my opinion should be included in any evaluation. Some of these have been suggested explicitly or implicitly in writings of Scriven (1967), Stake (1967, 1970) and Stufflebeam (1968), while some of the proposals originate with my views on evaluation. The result is a checklist of general characteristics of good evaluations which any school could use to determine whether its evaluation plan includes such important considerations.

1. Conceptual Clarity

Conceptual clarity, an essential feature of any good evaluation plan, refers to whether or not the evaluator exhibits a clear understanding of the particular evaluation he is proposing. Is he planning a formative or summative evaluation? Is it a comparative evaluation design or a single program evaluation? Is the evaluation to be goal-directed, with the design built around the measurement of attainment of specific objectives, or goal-free with the design built around lists of evaluative questions generated independently of the goals? Answers to questions such as these should be apparent in any good evaluation plan. Without clarity on these points, it would be an accident if the remainder of the evaluation were anything but a muddle.

2. Characterization of the Object of the Evaluation

No evaluation is complete unless it includes a thorough, detailed description of the program or phenomenon being evaluated. Without such characterization, judgments may be drawn about entities which never really existed.[9] For example, the concept of team teaching has fared poorly in several evaluations, resulting in a general impression that team teaching is ineffective. Closer inspection shows that many methods labeled as team teaching provided no real opportunities for staffs to plan together or work together in direct instruction. Obviously, a better description of the phenomenon would have avoided these misinterpretations completely. One simply cannot evaluate adequately that which he cannot describe accurately.

3. Recognition and Representation of Legitimate Audiences

An evaluation is adequate only if it includes input from and reporting to all legitimate audiences for the evaluation. An evaluation of a school program which answers only the questions of the school staff and ignores questions of parents, children, and community groups is simply a bad evaluation. Each legitimate audience must be identified and the evaluation plan should include their objectives or evaluative questions in determining what data must be collected. Obviously, some audiences will be more important than others and some weighting of their input might be necessary. Correspondingly, the evaluation plan should provide for receipt of appropriate evaluative information by each audience which has a direct interest in the program.

---

[9] Charters and Jones (1973) have claimed that such appraisal of "non-events" is frequent in program evaluation. However, their failure to present empirical evidence for their claims led Murray (1974) to waggishly suggest that critiquing the Charters-Jones paper might be evaluating a non-event.

4. Sensitivity to Political Problems in Evaluation

Many a good evaluation, unimpeachable in all technical details, has failed because of its political naiveté. It is pointless to promise to collect sensitive data (e.g., principals' ratings of teachers) without first obtaining permission from the office or individual who controls those data. Agreements must be reached early in any evaluation about issues such as access to data and data sources and safeguards against misuse of evaluation data. Steps must be taken to guarantee that program staff have opportunities to correct factual errors in evaluation reports without compromising the evaluation itself. These issues exist in almost every evaluation and the more explicitly they are dealt with, the more likely the evaluation is to survive political pressures.

5. Specification of Information Needs and Sources

Good evaluators tend to develop and follow a blueprint which tells them precisely what information they need to collect and what the sources of that information are. At the very least, they know how (as Scriven puts it) to lay snares at critical points in the game trails. Conversely, the novice evaluator goes about randomly turning over stones or beating the brush to see what he can find. No evaluation can depend on a random, scattered "here a little, there a little" approach to collecting data. An adequate evaluation plan specifies at the outset the information which must be collected. If the evaluation is goal-directed, the plan will specify information that will help to determine whether the objectives were attained. If the evaluation is built around evaluative questions (of the "What would you need to know to decide whether the program was a success or a failure?" variety), the evaluation plan should specify information which, when collected, will answer those questions. And in every case, listing of the needed information leads logically to identification of the sources from which that information can be

obtained. Failure to attend to these seemingly pedestrian but truly critical steps is one of the greatest single reasons so many evaluations produce little useful information.

6. Comprehensiveness/Inclusiveness

This category is really an elaboration of the previous one. No evaluation can hope to collect all of the relevant data, nor would it be desirable to do so, since there will always be inconsequential and trivial data not worth the bother to collect. Collecting too much data is seldom the concern, however. The greater problem is collecting enough data--or more precisely, collecting data on enough important variables to be certain one has included in the evaluation all the major considerations which are relevant. A good evaluation includes all of the main effects, but also includes provisions for remaining alert to unanticipated side effects. A good comparative evaluation doesn't stop with comparing the experimental arithmetic program with a control group which receives no arithmetic instruction. It goes on to identify the critical competitors--SMSG math, Cuisennaire Rods, and so forth--and compares their new program with those for which costs are roughly comparable. In short, the weak evaluation is almost always characterized by a narrow range of variables and omission of several which are important. The wider the range and the more important the variables included in the evaluation, the better it generally is.

7. Technical Adequacy

More evaluations fail here than on almost any other dimension, and this is due to the scarcity of educational evaluators who are even marginally competent in technical areas. Good evaluations are dependent on construction or selection of adequate instruments, the development of adequate sampling plans, and the correct choice and application of techniques for data reduction and

analysis. Volumes have been written on educational measurement, sampling, and statistics and it would be pointless to try to review that knowledge here. Suffice it to say that these areas are all essential to most evaluations. Without knowledge and control of these tools of his trade, the evaluator has little hope of producing evaluation information which meets scientific criteria of validity, reliability, and objectivity.

8. Consideration of Costs

Educators are not econometricians and should not be expected to be skilled in identifying all the financial, human, or time costs associated with programs they operate. That bit of leniency cannot be extended to the evaluator, however, for it is his job to bring these factors to the attention of developers, teachers and administrators who are responsible for their products or programs. Educators are often faulted for choosing the most expensive program from two that are equally effective, just because the expensive one is packaged more attractively or has been more widely advertised. The real fault lies with the evaluations of those programs which fail to consider cost factors along with the other variables. As any insightful administrator knows, costs are not irrelevant, and it is important for him to know how much Program X will accomplish and at what cost, so he may know what he is gaining or giving up in looking at other alternatives which range in both cost and effectiveness.

9. Explicit Standards/Criteria

It is always a bit disconcerting to read through an evaluation report and be unable to find anywhere a statement of the criteria or standards which were used to determine whether the program was a success or a failure. The

measurements and observations taken in an evaluation cannot be translated into judgments of worth without the application of standards or criteria. Is an in-service program for teachers successful if 75 percent of the teachers attend? That all depends on the rationale for the program and the attendance standard that would signal success or failure. What about a 70 percent attendance rate in a high school mathematics class--is that good or bad? Again, it depends on the standard. If it is a college preparatory class with high attendance expectations--say a standard of 95 percent--70 percent is very poor. If it is a remedial mathematics class for dropouts who are returning to school on a part-time basis, the expectation might be considerably lower--say 50 percent--and the attendance rate of 70 percent might be quite acceptable. These examples oversimplify the concept, but hopefully they will not detract from the point that every good evaluation will include a statement of standards and criteria.

10. Judgments and/or Recommendations

The only reason for insisting on explicit standards or criteria is that they are the stuff of which judgments and recommendations are made, and the latter are the sine qua non of evaluation. An evaluator's responsibility does not end with the collection, analysis, and reporting of data. The data do not speak for themselves. The evaluator who knows those data well is in the best position to apply the standards to the data to reach a judgment of whether the program is effective or ineffective, valuable or worthless. Making judgments and recommendations is an essential part of the evaluator's job. An evaluation without these ingredients is as much an indictment of its author's sophistication as one with recommendations that are not based on the data.

11. Reports Tailored to Audiences

It was maintained earlier that there are multiple audiences for most evaluations and these audiences have different information needs. For example, when an evaluator completes an evaluation, his methodologically oriented colleagues will be interested in a complete, detailed report of the data collection procedures, analysis techniques, and the like. Not so for the school board, or the PTA or the chairman of the local taxpayer group. These audiences do not share the evaluator's grasp of technical details, his interest in test reliability and validity, or his concern over the appropriate choice of an error term in a randomized blocks design. The evaluator will have to tailor reports for these groups so that they depend on non-technical language and avoid over-use of tabular presentation of data analyses. A typical evaluation might end up with one omnibus technical evaluation report which self-consciously includes all the details and one or more non-technical evaluation reports aimed at the important audience(s).

Another notion should be inserted here as well--that of interim or even continual reporting of evaluation findings. Timeliness is an important concern in evaluation. Information that is presented too late to affect the decision for which it is relevant is useless. Good evaluations will not depend solely on the printed word, but will include a variety of report formats, including "hot-line" telephone reporting, so the information is reported whenever it is needed to make a particular decision.

Conclusion

This paper was written for educators with little training or experience in formal evaluation of educational programs, products, or processes. In it,

I have attempted to provide a brief overview of a number of important considerations in educational evaluation and accountability. Specifically, the following topics have been presented: (1) simple illustrations of differences in evaluation, research, assessment, measurement, and accountability; (2) a discussion of some basic evaluation constructs; (3) an analysis of the evaluation features of one accountability law; and (4) general touchstones for judging the adequacy of an evaluation. Obviously, this sampling neglects many areas and results in an incomplete, oversimplified portrayal of the field. Hopefully, it will prove useful either as an evaluation primer for beginners or as a guide to sources of information for their further study. If any of the contents prove informative for more experienced evaluators, that would have to be viewed, in the current idion, as an "unanticipated side effect."

References

Charters, W. W., Jr., & Jones, J. E. On the risk of appraising non-events in program evaluation. Educational Researcher, 1973, 2 (11), 5-7.

Glass, G. V. The many faces of educational accountability. Phi Delta Kappan, 1972, 53, 636-639.

Glass, G. V, & Worthen, B. W. Educational evaluation and research: Similarities and differences. Curriculum Theory Network, Fall, 1971.

Murray, S. L. An untitled comment on Charters and Jones in the Educational Research and Development Evaluators Newsletter, 1974, 1(3), 2.

Scriven, M. The methodology of evaluation. In R. E. Stake (Ed.), Curriculum evaluation. American Educational Research Association monograph series on evaluation, no. 1, Chicago: Rand McNally, 1967.

Scriven, M. Prose and cons about goal-free evaluation. Evaluation Comment, 1972, 3(4), 1-4.

Stake, R. E. The countenance of educational evaluation. Teachers College Record, 1967, 68, 523-540.

Stake, R. E. Objectives, priorities, and other judgment data. Review of Educational Research, 1970, 40, 181-212.

Stufflebeam, D. L. Evaluation as enlightenment for decision-making. Columbus, Ohio: Evaluation Center, Ohio State University, 1968.

Stufflebeam, D. L., et al. Educational evaluation and decision-making in education. Itasca, Ill.: Peacock, 1971.

Wardrop, J. L. Towards a broader concept of educational assessment. Paper presented at the Second Annual Mardi Gras Symposium, New Orleans, February 1972.

Worthen, B. R. Impediments to the practice of educational evaluation. Research Paper No. 60. Boulder, Colo.: Laboratory of Educational Research, University of Colorado, 1972.

Worthen, B. R., & Sanders, J. R. Educational evaluation: Theory and practice. Worthington, Ohio: Charles A. Jones Publishing Co., 1973.

The following papers in the Research, Evaluation and Development Paper Series are available at the prices stated below:

- The Status of Research on Models of Product Development and Evaluation. By Nick L. Smith and Stephen L. Murray, Northwest Regional Educational Laboratory. $1.50 per copy (806-5201)

- Techniques and Procedures for Formative Evaluation. By James Sanders, Northwest Regional Educational Laboratory and Donald J. Cunningham, Indiana University. $3.00 per copy (806-5202)

- A Look at the Mosaic of Educational Evaluation and Accountability. By Blaine Worthen, Northwest Regional Educational Laboratory. $2.50 per copy (806-5203)

- The Administrator-Researcher Interaction: The Conduct of Cooperative Research. By Stephen L. Murray and Nick L. Smith, Northwest Regional Educational Laboratory. $1.50 per copy (806-5204)

- Pre-product Evaluation: What to do While Waiting for the Product. By Nick L. Smith, Northwest Regional Educational Laboratory. $1.50 per copy (806-5205)

- A Prescriptive Model of Development or Evaluation: Some Needed Maturity. By Cecil Clark, Brigham Young University. $2.50 per copy (806-5206)

- Some Tasks and Competencies Frequently Required in Educational Research and Evaluation. By Blaine Worthen, Northwest Regional Educational Laboratory. $2.50 per copy (806-5207)

- The Use of Multiple Strategies in the Evaluation of Experience-Based Career Education. By Thomas R. Owens, Joseph F. Haenn and Harry L. Fehrenbacher, Northwest Regional Educational Laboratory. $2.50 per copy (806-5208)

- The Use of Student Case Study Methodology in Program Evaluation. By Harry L. Fehrenbacher, Thomas R. Owens and Joseph F. Haenn, Northwest Regional Educational Laboratory. $2.50 per copy (806-5209)

Complete Set of Titles: $17.00 (reflects 15% discount)

Prices subject to change without notice.

Send purchase order or prepayment for any of the titles listed above to:

Office of Marketing and Dissemination
Northwest Regional Educational Laboratory
710 S.W. Second Avenue
Portland, Oregon 97204